

Niche Modeling of Dengue Fever Using Remotely Sensed Environmental Factors and BRT

Jeffrey L. Ashby* and Max J. Moreno-Madriñán

Fairbanks School of Public Health, Indiana University, Greencastle, IN, USA

Objective

In this paper we used Boosted Regression Tree analysis coupled with environmental factors gathered from satellite data, such as temperature, elevation, and precipitation, to model the niche of Dengue Fever (DF) in Colombia.

Introduction

Dengue Fever (DF) is a vector-borne disease of the flavivirus family carried by the *Aedes aegypti* mosquito, and one of the leading causes of illness and death in tropical regions of the world. Nearly 400 million people become infected each year, while roughly one-third of the world's population live in areas of risk. Dengue fever has been endemic to Colombia since the late 1970s and is a serious health problem for the country with over 36 million people at risk. We used the Magdalena watershed of central Colombia as the site for this study due to its natural separation from other geographical regions in the country, its wide range of climatic conditions, the fact that it includes the main urban centers in Colombia, and houses 80% of the country's population. Advances in the quality and types of remote sensing (RS) satellite imagery has made it possible to enhance or replace the field collection of environmental data such as precipitation, temperature, and land use, especially in remote areas of the world such as the mountainous areas of Colombia. We modeled the cases of DF by municipality with the environmental factors derived from the satellite data using boosted regression tree analysis. Boosted regression tree analysis (BRT), has proven useful in a wide range of studies, from predicting forest productivity to other vector-borne diseases such as Leishmaniasis, and Crimean-Congo hemorrhagic fever. Using this framework, we set out to determine what are the differences between using presence/absence and case counts of DF in this type of analysis?

Methods

We combined data on Dengue fever cases downloaded from the Instituto Nacional de Salud (INS) Programa SIVIGILA INS site with population data downloaded from the 2005 General Census administered by the National Administrative Department of Statistics (Departamento Administrativo Nacional de Estadística, DANE) and projected to 2012–2014 levels. We acquired remote sensing data from the National Aeronautics and Space Administration (NASA) data servers for each day of the study period. Imagery for each environmental variable was composited to reduce the effects of cloud cover and to match the ISO Week Date format reporting of the case data. We aggregated these weekly composite images for each variable using GIS to create annual minimum, maximum, and mean for a raster cell. These data were further aggregated to the municipality level using the GIS, again for minimum, maximum, and mean. Land use and elevation were only downloaded for one period given they change very little over time. The BRT analysis was conducted twice: once using the Bernoulli family of presence/absence and again using the Poisson family of actual case counts. In the first analysis (Bernoulli), any municipality reporting one or more cases of DF in the year was coded as having disease “presence”, while all others were coded as not having disease “absence”. The BRT model was run, using a twenty-five percent hold out of the data as a testing set, for each year. In the second analysis (Poisson), the only change to the models consisted of

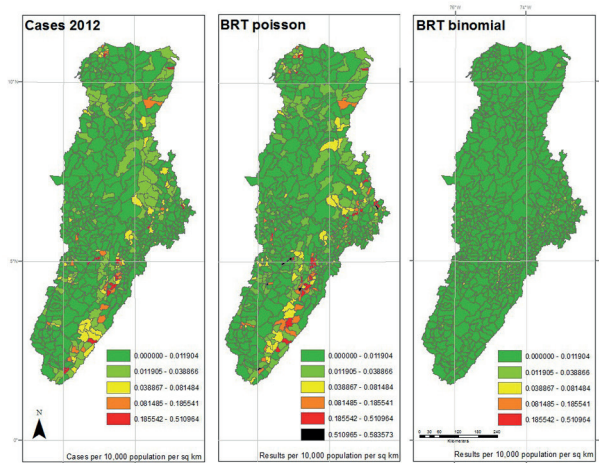
replacing the presence/absence data with the actual cases of reported DF within the municipality. The Poisson family was chosen in the model since the count data were highly skewed.

Results

We calculated RMSE and Pearson *r* values for each of the three years. The Poisson model out-performed the Bernoulli model across all years. The RMSE values were considerably lower for the Poisson model compared to the Bernoulli model, reflecting a better model fit. The Pearson *r* values were higher for the Poisson model compared to the Bernoulli model, again across all three years. We created maps to compare cases with the Poisson and the Bernoulli results. The maps shown in the figure reflect the results for 2012. The left panel represents the cases per 10,000 population per square kilometer for each municipality. The dark green color represents very low ratios of DF, while the red color reflects a higher incidence of DF. All maps used the same classification as the reported cases map for comparison, with an additional symbol (black) used for values outside the reported cases range.

Conclusions

Using actual reported case data and the Poisson function within the BRT functions created by Elith et al. and the *gbm* package in R, we show that the differences between using presence/absence and case counts of DF in a BRT analysis gives a clearer picture of the spatial distribution of DF. By using readily available and freely accessible data, we have shown that practitioners both within and outside of Colombia can quickly create accurate maps of annual DF incidence. The methods described here could also be extended to other regions and diseases, making it useful to a wide range of end users.



Maps showing the results for 2012

Keywords

Dengue; boosted regression tree; *Aedes aegypti*; remote sensing; GIS

*Jeffrey L. Ashby

E-mail: jlashby@iupui.edu

